



free procedures. In particular, nonparametric estimators of the failure rate,

IMCI ACCIPTED

THE COPPS

DD FORM 1473 EDITION OF I NOV AS IS OBSOLETE

20._(cont.)

function and survival probability were developed under the assumption of increasing failure rate using both maximum likelihood and Bayesian approaches. These particular results have attracted wide attention due to their generality and applicability in survival analysis and reliability estimation from arbitrarily right-censored data. Also, consistency results for both univariate and multivariate kernel estimates for probability density functions and regression functions were obtained using techniques and results of function-space probability theory. Sequential procedures were developed and analyzed which provided interval estimators of the parameter of interest after testing certain hypotheses. Various robustness and nonparametric methods for incomplete samples and broken samples were also studied. Thus, maintenance policies and development of new, more reliable, equipment may be formulated using statistical procedures and theory from these results.

Accession Fer	
NTIS GRAZI DDC TAB Unemnounced Justification	
By	
Availability Codes	
Dist.	Avail and/or special

Nonparametrice 8 0 70 5 7 1 Estimation of Distribution Functions and the Study of Probability Density Introduction Introduction

This document reports in detail the work performed and other research activities of the principal investigators during the period from June 1, 1979, to May 31, 1980, under contract number 49620-79-C-0140.

In Section 2 a comprehensive statement of the research objectives during this reporting period is given, and in Section 3 a detailed report of the progress made and highlights of a significant results obtained is presented. Section 4 contains a complete listing of approximately twenty-six publications which have resulted to date from these research efforts. The personnel supported under this contract are listed in Section 5, and a list of the professional interactions — talks at conferences and meetings, colloquia and seminars, etc. — is presented in Section 6. Section 7 gives an indication of the significance of the results to military application. Finally, other professional activities of the principal investigators are outlined in Section 8.

80 8 14 115

Approved for public release; distribution unlimited.

Research Objectives During the Reporting Period

The research objectives of this grant can be divided into four broad categories or areas:

- (A) Survival analysis and life testing;
- (B) Probability density estimates and laws of large numbers;
- (C) Estimation after testing;
- (D) Robustness and distribution-free procedures.

The research problems which were considered under this grant in these four areas will be outlined in this section. The progress toward their solution and significant results will be described in Section 3 of this report.

A. Survival Analysis and Life Testing

The main thrust of the research in this area was to develop estimators of survival probability or reliability in nonparametric settings or in a non-parametric Bayesian framework which were much better than existing estimators in the sense of their smoothness and convergence properties. The main objective in this area was to develop such estimators for arbitrarily right-censored observations under no distributional assumptions other than assumptions concerning the failure rate. In particular, since increasing failure rate (IFR) distributions are very commonly used to model wear-out failures, the distributional assumption was simply that the underlying life time distribution was IFR. One of the most promising approaches to these problems was to consider (non-parametric) maximum likelihood estimation of an increasing failure rate function r(t), $t \ge 0$, based on arbitrarily right-censored observations from the life distribution F and then to estimate the survival probability $\overline{F}(t) = 1 - F(t) = \exp(-\int_0^t r(u) du)$ from the maximum likelihood estimate of r. The other promising AIR FORCE GENICE OF SCIENTIFIC RESEARCH (AFSC)

exp(-\int_0^t :(u)du) from the maximum likelihood estimate of r. The other promising AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)

NOTICE OF THATACHITTIE TO DDC

This terminate of the been reviewed and is approved to the law AFR 190-12 (7b).

A. D. BLANK

Technical Information Officer

avenue of attack for these problems was to assume a prior distribution on the space of all increasing failure rate functions and obtain a nonparametric Bayes estimator of F or \overline{F} which was a continuous function. A type of Poisson "shock model" was successfully used as a prior distribution.

Another objective in this area was to develop powerful nonparametric tests for symmetry of a bivariate distribution function with observations which are subject to arbitrary right censorship. The usual sign test is not the best test to use in the case of censoring since a pair of observations, one of which is censored, may not be comparable.

Also, it was desired to obtain Bayes estimators of the reliability function for the inverse Gaussian distribution. This parametric model for life-time distributions has recently been proposed as an alternative to the lognormal model.

B. Probability Density Estimates and Laws of Large Numbers

This area of the research concentrated on obtaining consistency of probability density estimates and developing laws of large numbers which would have applications in the estimation of probability density functions. In particular, the kernel density estimation techniques were examined with the research objective of obtaining a wider class of kernel functions and more general bandwidth sequences which could be used while preserving the stochastic convergence of the estimate to the true underlying probability density function. The development of multivariate kernel density estimates with convergence properties which were similar to the univariate estimates was also a research objective.

In the kernel density estimation, a kernel function (usually a probability density function which the experimenter may choose) is centered at each observed

sample point and is scaled, depending on the sample size, so that its spread decreases and becomes more concentrated around the sample observation as the sample size increases. The estimate is then an average of these kernel functions which are centered at the sample observations. Hence, laws of large numbers in function spaces containing the possible kernel functions would yield consistency of these estimates and was a major research objective of the grant. Since the scaling changed as the sample size changed, the appropriate formulation involved weighted sums of arrays of function-valued random variables. Also, the possible different measures of dispersion of the kernel estimate $f_n(t)$ from the probability density function f(t) necessitated the consideration of laws of large numbers for Banach spaces and the function space D[0,1] as a research objective.

C. Estimation After Testing

Suppose F_{θ} is a distribution function which is stochastically nondecreasing in θ . It was desired to obtain various sequential procedures which allow quick rejection of the null hypothesis H_0 : $\theta \geq \theta_0$ when the alternative H_1 : $\theta \leq \theta_1$ ($\theta_0 > \theta_1$) is true. In addition, these procedures would provide an accurate interval estimate of the parameter θ when the null hypothesis is accepted. The results would hold in particular for an exponential distribution with mean θ .

n. Robustness and Distribution-free Procedures

Three of the research objectives were in this category. One is concerned with the distributional robustness when there are missing observations in the samples. The second one was to establish the robustness (asymptotic conservativeness) of a K-sample Kruskal-Wallis test under certain departures from mutual independence of K univariate samples. The third objective was to investigate an asymptotically distribution-free simultaneous confidence region of treatment differences in a randomized complete block design.

3. Status of the Research Effort

In this section, a substantive statement of the progress and significant accomplishments towards achieving the research objectives outlined in Section 2 of this report will be given. The specific research papers containing the results will be listed in Section 4.

A. Survival Analysis and Life Testing

One of the most significant results in this area is paper number 1 in the listing of Section 4. This paper has already received wide attention from workers in this area before it has appeared in print, since it was presented at the annual Institute of Mathematical Statistics Meeting in Washington, D. C., in August, 1979. This result concerns the (nonparametric) maximum likelihood estimation of an increasing failure rate function and the corresponding distribution function based on a set of observations subject to arbitrary right censorship. The estimator is developed using the techniques of isotonic regression, is defined everywhere on the positive real line, is a smooth function, and decreases to zero as time increases. The small sample properties of the estimator were compared with the Kaplan-Meier estimator by a Monte Carlo study for Weibull distributions, and the results indicated that the maximum likelihood estimator is superior to the Kaplan-Meier estimator. This estimator is applicable to a broad spectrum of problems due to the generality of the censoring mechanism. For example, medical trial data as well as industrial life test data may be analyzed with this estimation procedure. In particular, since no assumptions are required concerning the exact lifetime distribution, this estimator is especially applicable to the assessment of military equipment reliability and maintenance. Such equipment which is subject to wearout may be assumed to follow

an IFR life distribution, and this is the only assumption necessary for the estimation procedure developed here.

A second significant result in this area is contained in paper number 2 in Section 4. Bayesian nonparametric estimators of the survival function, the failure rate function, and the density function are obtained using Poisson jump processes to generate prior probability measures on the space of increasing failure rate functions. The jump processes are intuitively appealing and have a meaning physical interpretation as a shock model, where "shocks" occurring to the system increase the failure rate by a positive amount at the random instant of occurrence. The Bayesian nonparametric estimators are also easily derived for these priors for arbitrarily right-censored observations. These estimators are continuous and are thus more appealing in this sense than estimators of other authors who used Dirichlet processes as prior distributions for (discrete with probability one) distributions. The same comments made on the applicability of the estimator in the last paragraph can be made for these nonparametric Bayes estimators. Also, along these same lines when the Dirichlet process is not completely specified, several sequences of nonparametric empirical Bayes estimators of survival probability were considered (paper number 18 in Section 4) and their asymptotic optimality with respect to a Dirichlet process was investigated.

In testing a bivariate life distribution for symmetry based on arbitrarily right-censored observations, a conditionally distribution-free test was developed (paper number 6 in Section 4). A permutation test using Gehan's scores is proposed which is much more powerful than the sign test. The power of the test was compared with that of the sign test by computer simulations using the Marshall-Olkin bivariate exponential model.

For the inverse Gaussian distribution as a parametric lifetime model, Bayes estimators of the reliability function have been obtained (paper number 13 in Section 4). For the case that the mean lifetime is known, Bayes estimators of the reliability are obtained with respect to squared error loss for Jeffreys' noninformative prior and for the gamma family (natural conjugate family) of priors for the scale parameter. In the case that both the mean and scale parameter are unknown, an estimator of reliability is suggested which is based on the Bayes estimator for the first case. This estimator compares favorably with the maximum likelihood and minimum variance unbiased estimators.

B. Probability Density Estimators and Laws of Large Numbers

In paper number 3 in the listing of Section 4, the uniform consistency of the multivariate kernel density estimate, $\hat{f}_n(x_1, \dots, x_k)$, was obtained by showing that $\|\hat{f}_n(x_1, \dots, x_k) - f(x_1, \dots, x_k)\|_{\infty}$ converges completely to zero for a large class of weight functions and under mild conditions on the bandwidth sequences $b_k(n)$'s. This significant result provided for the uniform consistency of the estimate $\hat{m}_n(x_1, \dots, x_k) = \frac{\hat{h}_n(x_1, \dots, x_k)}{\hat{f}_n(x_1, \dots, x_k)}$ for the regression function

 $m(x_1,\ldots,x_k)=E(Y|X_1=x_1,\ldots,X_k=x_k)$ where $f(x_1,\ldots,x_k)=\int_{f}^{*}(x_1,\ldots,x_k,y)dy$ and $h(x_1,\ldots,x_k)=\int_{f}^{*}(x_1,\ldots,x_k,y)dy$. In these results the rates of convergence need not be uniform in each coordinate and the broad class of possible kernels included Epanechnikov's optimal kernel and was characterized by a smoothness condition.

In papers number 8 and 9 in the listing of Section 4, the stochastic convergence of weighted sums of arrays of random elements in Banach spaces was obtained under varying conditions. One striking result was that the convergence of these weighted sums was necessary and sufficient for the geometric property,

type p, of the Banach space. As corollaries, these results have Banach space versions of several random variables results. In paper number 9, it is shown in the application of these results that the consistency in the L¹-norm of certain density estimates is sufficiently implied by $\sum_{n=1}^{\infty} (b_n^2 n)^{-q} < \infty$ for some q > 1 where n is the sample size and b is the bandwidth. The L^p-consistency for a broad class of kernel estimates is obtained from the results of paper number 8 when $\sum_{n=1}^{\infty} (nb_n)^{-q} < \infty$ for some q > 0. Hence, the bandwidth sequence b_n could be of order $n^{-\delta}$ for any $0 < \delta < 1$. This significant improvement for general classes of kernel functions and density functions (including multivariate considerations) is mainly due to the probability cancellation in type p spaces and the L^p norm.

In papers numbers 7 and 17 laws of large numbers for D[0,1] were obtained. These results considerably strengthened the existing results and provided much less restrictive hypotheses. The invitation to present related work at the International Conference on Probability in Banach Spaces is evidence of the significance of these results.

C. Estimation After Testing

In the quality control of industrial or military production, it is common practice to test hypotheses to make a decision about acceptance or rejection of a production batch. Generally, a sequential testing procedure requires less data to reach a decision than a fixed sample or fixed time testing procedure. Also, in many situations, the experimenter may want to obtain a point or an interval estimate of some parameter after a test of hypothesis has been established.

Let F_θ be a distribution function which is stochastically nondecreasing in $\theta.$ In particular, F_θ can be an exponential distribution with mean $\theta.$

Several sequential procedures were developed which allow quick rejection of the null hypotheses $H_o\colon\theta\geq\theta_o$ when the alternative $H_1\colon\theta\leq\theta_1(\theta_1<\theta_o)$ is true under the general model F_θ . In addition, these results provide accurate interval estimators of the parameter θ after the test has been completed (papers 10,11,20).

D. Robustness and Distribution-free Procedures

Firstly, a simple and robust estimator of the difference of location parameters of correlated variables is proposed when some observations on either of the variables are missing. We show that this estimator is consistent, asymptotically normally distributed, and insensitive to outlying observations. Asymptotic relative efficiency comparisons with other known estimators are made to show the advantage of the proposed estimator (paper 15).

Secondly, the robustness (asymptotic conservativeness) of Kruskal-Wallis test under certain departures from mutual independence of K univariate samples is established (paper 19). This robustness provides a procedure for testing the equality of K marginal distribution functions based on a <u>broken</u> random sample from a K-variate distribution which satisfies a mild condition. For the <u>unbroken</u> sample, a generalized Kruskal-Wallis test is proposed for testing the symmetry of a K-dimensional distribution function. The relative efficiency of the K-W test against the aligned rank order test is also examined under the normal shift model.

Thirdly, for a randomized complete block design with additive block effects, an asymptotically distribution-free simultaneous confidence region of pairwise treatment differences is presented. The corresponding confidence bound has an explicit form and is easily obtained. An example is provided for illustration purpose. The case of treatment against control is also discussed (papers 14 and 16).

4. Cumulative List of Written Publications in Technical Journals

In this section, the research papers which have been written under this grant are listed. They are divided into three categories: Accepted for publication, submitted, and manuscripts in preparation. Copies of papers 1-20 have already been forwarded to the Program Manager as they were submitted for publication.

A. Papers Accepted for Publication

- W. J. Padgett and L. J. Wei, Maximum likelihood estimation of a distribution function with increasing failure rate based on censored observations, (revised Univ. of S. C. Statistics T.R. #39), to appear in Biometrika (1980).
- 2. W. J. Padgett and L. J. Wei, A Bayesian nonparametric estimator of survival probability assuming increasing failure rate, (revised Univ. of S. C. Statistics T. R. #42), to appear in <u>Communications in Statistics</u> Theory and Methods (1980).
- 3. R. L. Taylor and K. F. Cheng, On the uniform complete convergence of estimates for multivariate density functions and regression curves, (Univ. of S. C. Statistics T. R. #34R), to appear in Annals of the Institute of Statistical Mathematics 32 (1980).
- 4. L. J. Wei, A generalized Gehan and Gilbert test for paired observations which are subject to arbitrary right censorship, to appear in <u>Journal</u> of the American Statistical Association (1980).
- 5. L. J. Wei, Recent development in the design of sequential medical trials, Proc. Conf. on the Recent Developments in Statistics, Taipei, Taiwan, to appear.

B. Papers Submitted for Publication

- 6. W. J. Padgett and L. J. Wei, Conditionally distribution-free test for censored bivariate observations, (Univ. of S. C. Statistics T. R. #41), submitted to Communications in Statistics Theory and Methods.
- 7. R. L. Taylor and P. Z. Daffer, On the weak law of large numbers for D[0,1], (Univ. of S. C. Statistics T. R. #38R), submitted to Zeitschrift für Wahrscheinlichkeitstheorie.

- 8. R. L. Taylor, Convergence of weighted sums of random elements in type p spaces, (Univ. of S. C. Statistics T. R. #43), submitted to Zeitschrift für Wahrscheinlichkeitstheorie.
- 9. R. L. Taylor, Complete convergence for weighted sums of arra s of random elements, (Univ. of S. C. Statistics T. R. #48), submitted to <u>Pacific</u> Journal of Mathematics.
- 10. W. J. Padgett and L. J. Wei, A sequential test and interval estimation in time truncated life testing, (Univ. of S. C. Statistics T. R. #49), submitted to Sankhya.
- 11. W. J. Padgett and L. J. Wei, Interval estimation after sequential testing based on the total time on test, (Univ. of S. C. Statistics T. R. #50), submitted to Naval Research Logistics Quarterly.
- 12. R. L. Taylor and P. Z. Daffer, Tightness of measures and laws of large numbers for generalized random variables, (Univ. of S. C. Statistics T. R. #35R), submitted to Journal of the Australian Mathematical Society.
- 13. W. J. Padgett, Some Bayes estimators of reliability for the inverse Gaussian lifetime model, (Univ. of S. C. Statistics T. R. #53), submitted to Technometrics.
- 14. L. J. Wei, The analysis of a randomized complete block design when observations are subject to arbitrary right censorship, submitted to Scandinavian Journal of Statistics.
- 15. L. J. Wei, A robust estimator of the difference between location parameters of correlated variables for fragmentary samples, submitted to Biometrika.
- 16. L. J. Wei, Asymptotically distribution-free simultaneous confidence region of treatment differences in a randomized complete block design, submitted to Journal of American Statistical Association.
- 17. R. L. Taylor and C. A. Calhoun, On the almost sure convergence of weighted sums of random elements in D[0,1], (Univ. of S. C. Statistics T. R. #56), submitted to the <u>International Journal of Mathematics and Mathematical</u> Science.
- 18. K. Y. Liang and W. J. Padgett, Nonparametric empirical Bayes estimation of reliability, (Univ. of S. C. Statistics T. R. #58), submitted to Metrika.
- 19. L. J. Wei, Asymptotic conservativeness and efficiency of Kruskal-Wallis test for k dependent samples, submitted to <u>Journal of American Statistical Association</u>.

C. Papers in Preparation May 31, 1980

20. W. J. Padgett and L. J. Wei, Interval estimation after sequential testing for the mean of the exponential distribution in the large sample case, (in revision).

- 21. L. J. Wei, The estimation of decreasing residual lifetime, (in revision).
- 22. L. J. Wei, The random missingship, (in progress).
- 23. R. L. Taylor, J. Howell, and W. Woyczynski, Stability of linear and quadratic forms for independent random variables in Banach spaces, to be submitted to <u>Annals of Probability</u>.
- 24. R. L. Taylor and P. Z. Daffer, Some strong and weak laws of large numbers in D[0,1], to be submitted to <u>Proceedings Third International</u> Conference on Probability in Banach Spaces, 1980.
- 25. R. L. Taylor and J. Howell, Marcinkiewicz-Zygmund type weak laws of large numbers for unconditional basic random elements in Banach spaces, to be submitted to <u>Proceedings Third International Conference on Probability in Banach Spaces</u>, 1980.
- 26. W. J. Padgett and L. J. Wei, Estimation of the ratio of scale parameters in the two sample problem with arbitrary right censorship, to be submitted to Journal of American Statistical Association.

5. Professional Personnel Associated with the Research Effort

In addition to the principal investigators, Professors Padgett,
Taylor, and Wei, D. T. McNichols is a research assistant under the grant and
has performed the computing for many examples used in the research papers,
performed reference searches, and proofread manuscripts. This student successfully completed the written and oral comprehensive examinations for the Ph.D.
program in September, 1979, and is beginning to embark into an area of research
for the Ph.D. dissertation.

Additionally, during the grant period, Professor Taylor has directed two students for their degrees of Master of Science: (i) Patrick E. Flanagan, May, 1980, "Computer programs for kernel estimates of a probability density function"; (ii) Carol A. Calhoun, May, 1980, "Laws of large numbers for D[0,1] and estimation of density functions." Professor Padgett has also directed Y. S. Lin for the degree of Master of Science. The title of his thesis is "Confidence bounds on reliability for the inverse Gaussian model," completed in May, 1980.

6. Interactions

The principal investigators presented (invited and contributed) papers at various conferences.

- (i) W. J. Padgett and L. J. Wei, "The maximum likelihood estimation of a distribution function with monotone failure rate based on censored observations." The Annual Joint Meetings of ASA and IMS in Washington, D.C.
- (ii) R. L. Taylor (joint with P. Z. Daffer), "Convergence of weighted sums of random elements in D[0,1]. Annual Meeting of IMS in Washington, D.C.
- (iii) L. J. Wei, "A generalized Gehan and Gilbert test for paired observations which are subject to arbitrary right censorship." Annual Meeting of ASA in Washington, D. C.
 - (iv) L. J. Wei, "The recent development of designs of sequential medical trials" (invited talk). Conference on Recent Development of Statistics at Academia Sinica, Taiwan.
 - (v) W. J. Padgett (joint with L. J. Wei), "A nonparametric Bayesian estimator of survival probability assuming increasing failure rate". Eastern Regional Meeting of IMS, Charleston, S. C.
 - (vi) L. J. Wei, "Analysis of randomized complete block design with arbitrary right censorship" (invited talk). Eastern Regional Meeting of IMS, Charleston, S. C.
- (vii) W. J. Padgett and Y. S. Lin, "Some confidence bounds on reliability for the inverse Gaussian model". The 2nd Annual meeting of S. C. Chapter of ASA.
- (viii) R. L. Taylor, "Complete convergence for weighted sums of arrays of random elements." Eastern Regional Meeting of the Institute of Mathematical Statistics, March 12-14, 1980, Charleston, S. C.
 - (13) R. L. Taylor and Carol Calhoun, "Strong laws of large numbers for D[0,1]". The 2nd Annual Meeting of the S. C. Chapter of ASA.
 - (x) L. J. Wei, "Asymptotically distribution-free simultaneous confidence region of treatment differences in a randomized complete block design". The 2nd Annual Meeting of the S. C. Chapter of ASA.

In addition, Padgett gave a colloquium talk at Auburn University.

Wei gave talks at National Cancer Institute, University of Maryland-College Park,

University of North Carolina-Chapel Hill, University of Wisconsin-Madison, and

Ohio State University.

7. Inventions, Patent Disclosures, and Applications Stemming from the Research Project

No inventions or patents have stemmed from this research.

The results reported in Sections 3 and 4 have wide application in the estimation and assessment of reliability of military equipment. Thus, maintenance policies and development of new more reliable equipment may be formulated using statistical procedures and theory from our general results. In particular, for reliability and survival analysis, accurate and convenient estimates of survival probabilities, probability density functions, and failure rate functions can be obtained only assuming an increasing failure rate. Hence, assumptions of various parametric failure models are not necessary for estimation for different types of equipment.

8. Other Professional Activities

During this reporting period, the principal investigators have been involved in numerous professional activities which are intimately related to the research efforts on this grant. The principal investigators have refereed a total of eight manuscripts during June 1, 1979 to May 31, 1980, two for the Journal of Multivariate Analysis, one for Glasnik Matematicki, one for IEEE Transactions on Reliability, two for the Journal of the American Statistical Association, one for Biometrics, and one for the Journal of Statistical Inference and Planning. In addition, four reviews of papers were written for Mathematical Reviews and three for the Zentralblatt für Mathematik.

W. J. Padgett and R. L. Taylor were the program co-chairmen for the Eastern Regional Meeting of the Institute of Mathematical Statistics held on March 12-14, 1980. They were responsible for arranging six invited speaker sessions for this joint meeting with ENAR of the Biometric Society and certain sections of the American Statistical Association. Also, R. L. Taylor was chairman of a session on "Reliability and Growth Models" at the 1980 SREB Summer Research Conference on Statistics in Pensacola, Florida, and has been invited to talk at the Third International Conference on Probability in Banach Spaces to be held August 4-15, 1980, in Boston, Massachusetts.